Manchester Metropolitan University

Unal, Yurdagul and Chowdhury, Gobinda and Kurbanoğlu, Serap and Boustany, Joumana and Walton, Geoffrey (2019)*Research data management and data sharing behaviour of university researchers'.* In: ISIC: the information behaviour conference, 09 October 2018 - 11 October 2019, Krakow, Poland.

https://e-space.mmu.ac.uk

Research data management and data sharing behaviour of university researchers

Yurdagül Ünal, Gobinda Chowdhury, Serap Kurbanoğlu, Joumana Boustany, and Geoff Walton

Introduction. *The aim of this study is to understand how university researchers behave in the context of using and sharing research data in OA mode.*

Method. *An online questionnaire survey was conducted amongst academics and researchers in three countries – UK, France and Turkey. There were 26 questions to collect data on: researcher information, e.g. discipline, gender and experience; data sharing practices, concerns; familiarity with data management practices; and policies/challenges including knowledge of metadata and training.*

Analysis. *SPSS was used to analyse the dataset, and Chi-Square tests, at 0.05 significance level, were conducted to find out association between researchers' behaviour in data sharing and different areas of research data management (RDM).*

Findings. *Findings show that OA is still not common amongst researchers. Data ethics and legal issues appear to be the most significant concerns for researchers. Most researchers have not received any training in RDM such as data management planning metadata, or file naming. However, most researchers would welcome formal training in different aspects of RDM.*

Conclusion. *This study indicates directions for further research to understand the disciplinary differences in researchers' data access and management behaviour so that appropriate training and advocacy programmes can be developed to promote OA to research data.*

Introduction

Data, often described as the '*glue of a collaboration*' ([Borgman, Wallis and Mayernik, 2012](#), p. 485) and the '*lifeblood of research*' ([Borgman, 2012](#), p. 1066), transcends all domains of scholarship, and includes

factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. ([OECD, 2007](#), p. 13)

Other researchers have expressed a similar view (see for example, [Uhlir and Cohen, 2011](#)). The wealth of data, now available to researchers, has been described as a '*deluge of data*' ([Frank and Pharo, 2016](#); [Borgman, Wallis and Enyedy, 2007](#); [Borgman, 2015](#)). Sharing and re-use of research data can be immensely beneficial; e.g., it is estimated that,

the $13 billion in government spending on the Human Genome project and its successors has yielded a total economic benefit of about $1 trillion. A British study of its public economic and social research database found that for every £1 invested by the government, an economic return of £5.40 resulted ([The data harvest, 2014](#)).

Similar views have been expressed by other researchers ([Beagrie and Houghton, 2013](#); [Ball and Duke, 2015](#); [Borgman, 2012](#)). Government and research funding bodies therefore strongly advocate open access (OA) to research data. *The concordat on open research data* ([2015](#), p. 2) agreed by UK research funding bodies remark that '*the societal benefits from making research data open are potentially very significant; including economic growth, increased resource efficiency, securing public support for research funding and increasing public trust in research'*. However, the take-up of OA data amongst researchers has been low for several reasons ([The data harvest..., 2014](#); [Faniel, Kansa, Whitcher Kansa, Barrera-Gomez and Yakel, 2013](#); [Faniel, Kriesberg and Yakel, 2012](#); [Yakel and Faniel, 2014](#)).

Proper management of data throughout the research process is crucial for making it openly accessible, intelligible, assessable and usable. The European Commission has developed a set of guidelines for data management in order to help researchers make their research data findable, accessible, interoperable and reusable (FAIR) ([European Commission, 2016](#)). Principle 3 of *The concordat on open research data* (*[2015](#)*, p.5) makes specific recommendations for searchability and usability of research datasets: '*In order to make data open, it must be not only accessible, but also discoverable and useable'*. The *Research Data Primer*, developed by the US National Information Standards Organization (NISO) ([Strasser, 2015](#)), suggests that a Data Management Plan (DMP) should include a description of all types of data, a description of all types of metadata and policies used, plans for archiving and preservation, and a description of resources required for data management.

However, little is known about how researchers manage and share their data with others outside of the immediate research collaboration (Borgman, 2012; Mayernik, Wallis and Borgman, 2013; Wallis, Rolando and Borgman, 2013). Within the '*long tail*' of small research projects data sharing is described as a '*gift culture*' (Wallis *et al.*, 2013) where data is bartered between colleagues in trusted relationships; however, the further someone is from this type of '*long-tail*' original research the harder it is to make use of this data (Borgman, 2012).

Although data sharing is regarded as a good behaviour, it appears that researchers invest little time in metadata and documentation rendering a great deal of data undiscoverable (Borgman, 2012; Wallis *et al.*, 2013). Without contextual information, where data has been separated from context, reuse can become '*difficult or impossible*' (Koltay, 2015, p. 405). In fact, according to MacMillan (2014), very few researchers (22%) use metadata, preferring to use their own laboratory standards instead. This view supported by Carlson, Fosmire, Miller and Nelson (2011).

Research questions and methods

The aim of this ongoing study is to understand how university researchers behave in the context of using and sharing research data with other researchers, and with everyone in OA mode. More specifically the study reported in this paper aims to address the following questions:

- What are the typical behaviours of university researchers in the context of data sharing and open access to data?
- What are the typical behaviours of researchers with regard to research data management?
- What are the training needs and requirements regarding research data management?

An online questionnaire survey was conducted amongst academics and researchers in three countries – UK, France and Turkey. These countries were chosen purposely to represent examples to see whether the different levels of development nationally with regard to research data management policies and practices, have any associations with researchers' data management practices and sharing behaviour. Amongst these three countries, UK is arguably the most advanced in terms of research and development of technologies, tools and policies for RDM. Researchers led by agencies like JISC (https://www.jisc.ac.uk) and DCC (https://www.dcc.ac.uk), and specific universities such as Oxford, Edinburgh, UCL and Southampton, have been engaged in research in different areas of RDM for nearly a decade. A significant move towards management and research data sharing is also evident through various

policies recently introduced by government funding agencies in UK; for example, the RCUK (Research Councils UK) *Common principles on data policy* ([Research Councils UK, 2015](#)).

In France, interest in research data was first stated publicly in 2011 when the Ministry of Higher Education and Research implemented a platform for monitoring and providing information about research data, and for raising awareness and encouraging a debate around challenges related to research data (http://www.donneesdelarecherche.fr/). Two years later, still under the aegis of the same Ministry, a research infrastructure *Huma-Num* was created '*to facilitate the digital turn in humanities and social sciences… with a range of utilities to facilitate the processing, access, storage and interoperability of various types of digital data*'. (http://www.huma-num.fr/about-us). Since then many academic libraries created specific services on RDM, e.g. Paris Descartes University in collaboration with Paris 7 published a guide on how to elaborate a data management plan ([Cartier, Moysan and Reymonet, 2015](#)).

In Turkey, neither the national funding agency for scientific research (TUBITAK) nor other funding agencies or universities have any RDM policy and/or mandate. There are no units within research institutions which provide support to researchers for storage and sharing of research data ([Tonta, 2013](#); [Aydınoğlu, 2016](#)). However, there have been several initiatives for raising awareness in data management including symposia ([Aydınoğlu, Doğan and Taşkın, 2017](#)), workshops (http://rdm.bilgiyonetimi.net/index.html), and conferences ([Aydınoğlu, 2016](#); [Gürdal and Bitri, 2015](#); [Aydınoğlu, 2014](#); [Tonta, 2013](#)).

Given the varying levels of technology and policy developments for RDM in the UK, France and Turkey, it was envisaged that a study in these countries would enable an understanding of the level of awareness of, and preparedness for, RDM amongst academics and researchers to emerge. Furthermore, it would help us explore the association (if there is any) between the current state of RDM developments and policies in the country and the researchers' RDM behaviour and practices.

Researchers' disciplines were grouped in three broad categories to make the computation of associations easy:

- Sciences (natural sciences, engineering and technology, medical and health sciences, agricultural sciences)
- Social sciences (for example education, sociology, law and political science)
- Humanities (for instance history, archaeology, language and literature)

E-mail invitations requesting for voluntary participation, and some subsequent reminders, were sent out using mailing lists for university academics and researchers with a request to fill-out an online questionnaire. There were 26 questions to collect data on: researcher information, e.g. role, discipline, gender and experience; data sharing practices, concerns; familiarity with data management practices; and policies/challenges including knowledge of metadata and training (See: https://inlitas.org/survey/index.php/493475?newtest=Y&lang=en). SPSS was used to code and analyse the dataset, and Chi-Square tests, at 0.05 significance level, were conducted to find out association between researchers' behaviour in data sharing as well as different areas of research data management and their demographic characteristics such as country, discipline and years of experience. Any associations detected with such demographic features could be useful for practitioners to develop training and advocacy programs accordingly.

Findings

Study population

The survey was conducted over several months in 2016-2017, and a total of 1098 complete responses were received. Out of 1098, 49% of the participants were from Turkey, 41% from France and 10% from UK; two-thirds of the respondents (76%) were academic staff; 58% of respondents were from sciences, 29% from social sciences, and 13% from humanities; 7% of participants had less than 5 years of experience, 24% had 5-10 years' experience, 15% had 11-15 years' experience, 13% had 16-29 years' experience, and 24% had over 20 years' experience.

Collaboration, data sharing and open access practices

Most researchers claimed that they collaborate and share data with others in one way or another: 73% collaborate and share data with researchers in the same team, 42% collaborate and share data with researchers in the same university, and 55% collaborate and share data with researchers in other institutions (Figures 1-3).
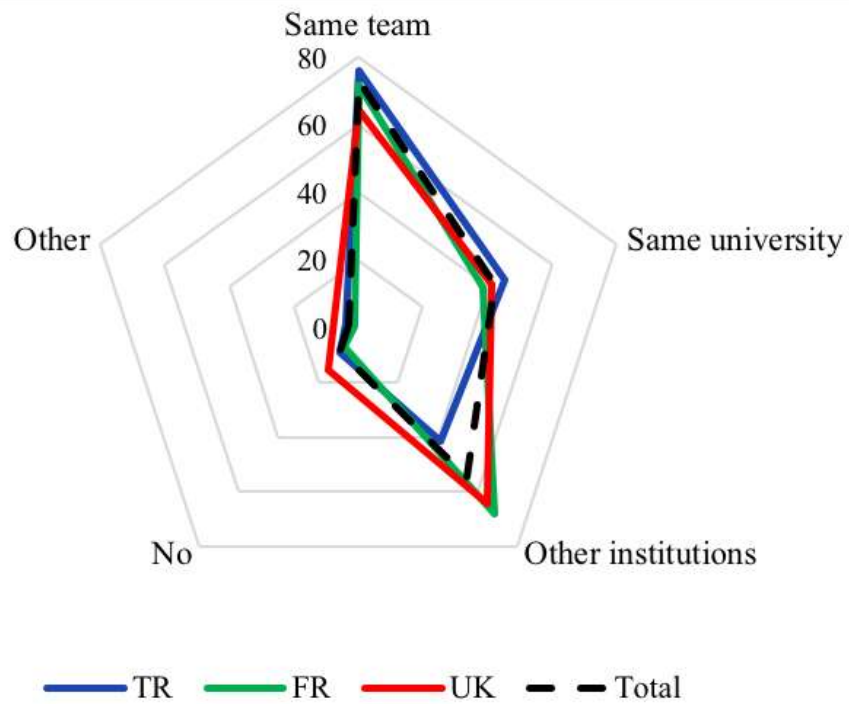
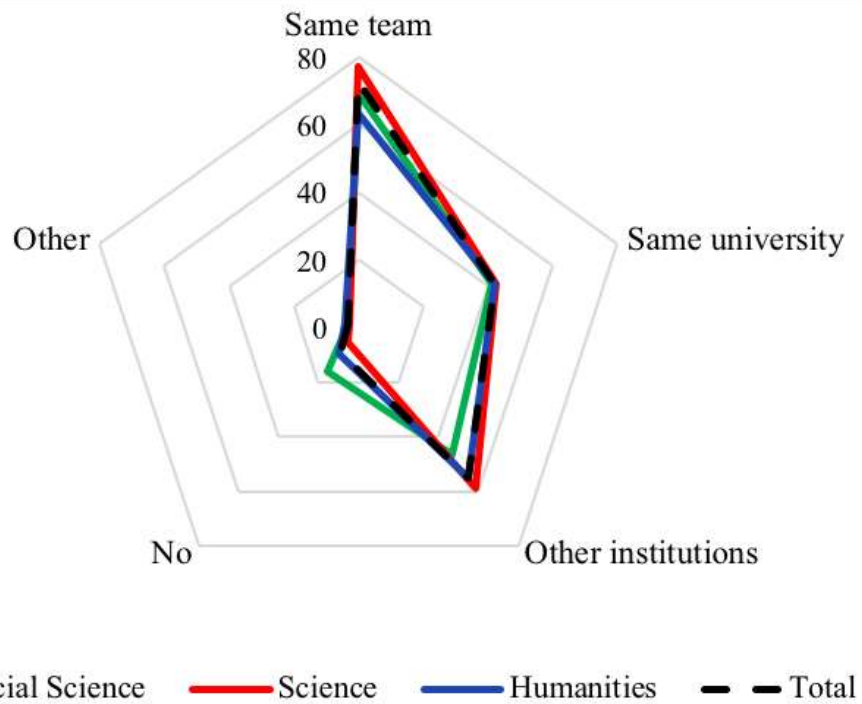Figure 1: Collaboration and data sharing by country

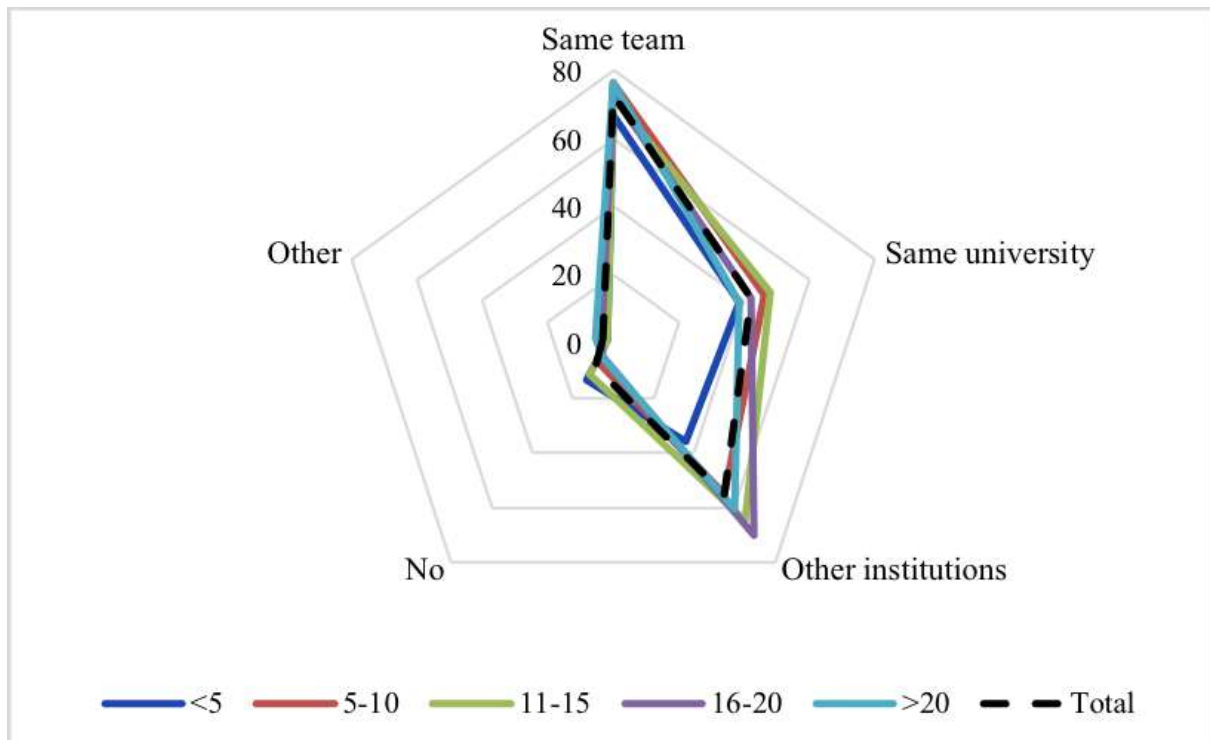Figure 2: Collaboration and data sharing by discipline

Figure 3: Collaboration and data sharing by experience

Statistically significant associations were detected between collaboration and data sharing behaviours and researchers' attributes such as country, discipline, and years of experience (Table 1). Findings indicate that collaboration and data sharing with researchers from other institutions is highest in France (68%) and lowest in Turkey (42%); researchers in science disciplines are more open to collaboration and data sharing than the others. There is an association with years of experience and degree of collaboration, those researchers who are least experienced tend to collaborate the least. A researcher's network tends to expand over time and this could explain this particular finding.

Table 1: Associations between researchers' attributes and collaboration and data sharing		
Collaboration and data sharing		**Value**
Researchers' country	no collaboration and data sharing	$X^2_{(2)} = 7.377$; p = 0.025
	with researchers in the same team	$X^2_{(2)} = 7.006$; p = 0.030
	with researchers in other institutions	$X^2_{(2)} =$ **75.854**; p = 0.000
Researchers' discipline	no collaboration and data sharing	$X^2_{(2)} = 29.566$; p = 0.000
	with researchers in the same team	$X^2_{(2)} = 14.526$; p = 0.001

	with researchers in other institutions	$X^2_{(2)}$=12.227; p = 0.002
Researchers' years of experience	no collaboration and data sharing	$X^2_{(5)}$=19.500; p = 0.002
	with researchers in other institutions	$X^2_{(5)}$=**64.875**; p = 0.000

Forty three percent of the researchers *almost always* or *often* (15% and 28% respectively) use OA datasets; and the figure decreases if the data has restricted access (11% *almost always*, 18% *often*); 86% of researchers need to spend some time to pre-process data received from outside sources: in 51% cases *a bit of effort* is required, while in 36% cases *a lot of effort* is required.

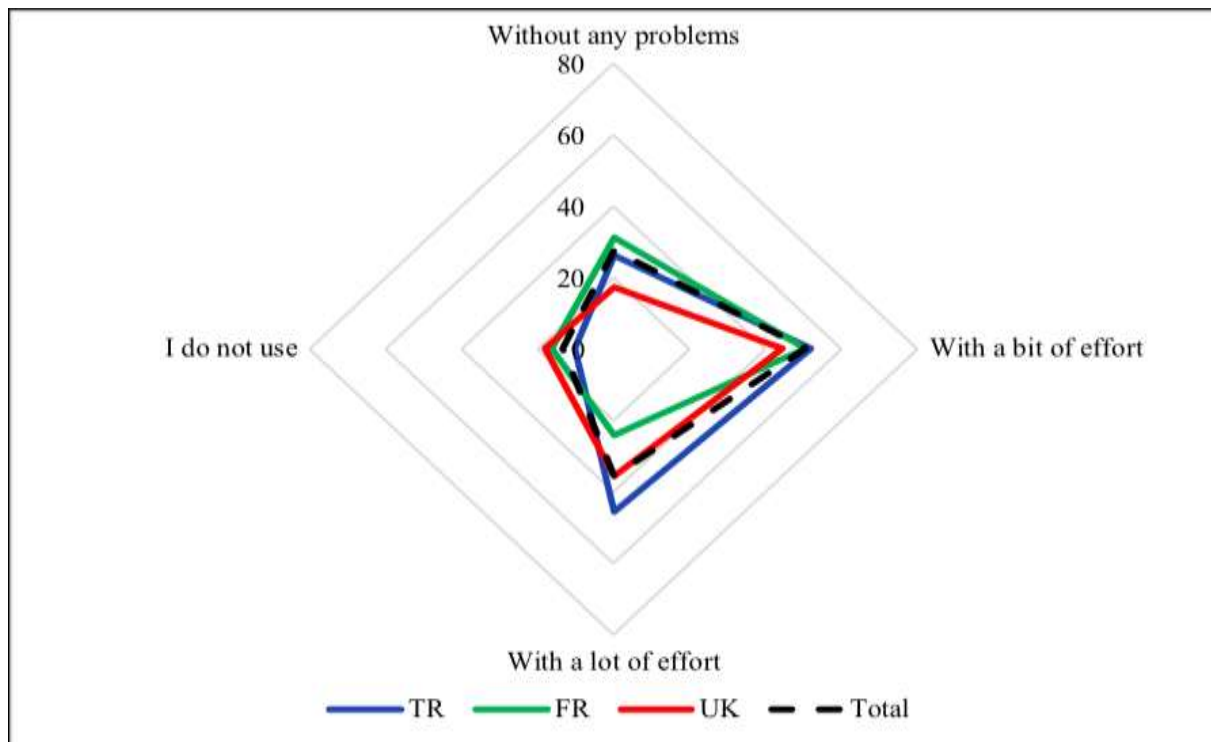About 14% of researchers do not use data from others/outside sources (Figures 4-5).

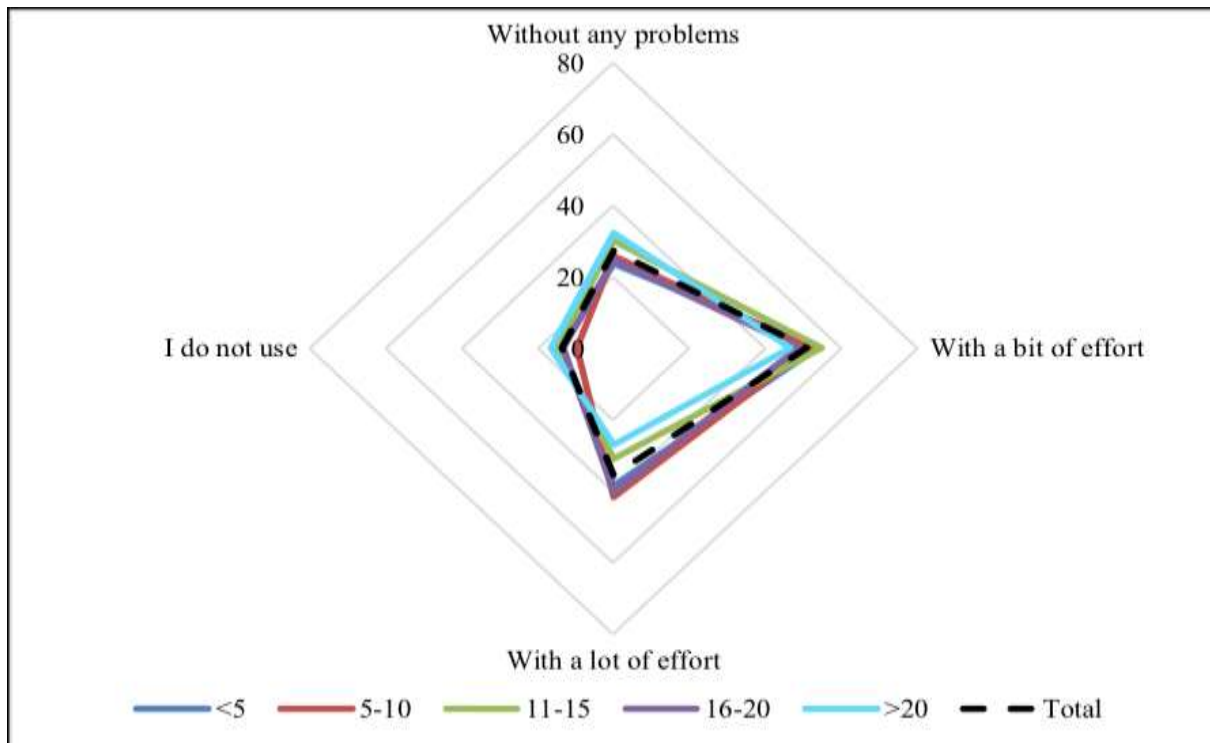Figure 4: Use of data from outside sources by country

Figure 5: Use of data from outside sources by discipline

Associations were noted between researchers' attributes and some of their behaviours in relation to the pre-processing of research data obtained from outside sources (Table 2). For instance, there is a statistically significant association between the effort spent on the data from outside sources and country. Almost half of researchers (46%) from Turkey spend *a lot of time and effort* to pre-process data received from outside sources (as opposed to 36% from UK and 24% from France). These findings could be linked with the development of RDM related practices, policies and mandates in the respective countries. The lowest effort for pre-processing is required by the most experienced group (>20 years' experience: 27%). This could be linked with the well-established and trustworthy research network they might have.

Table 2: Associations between researchers' attributes and use of data from outside sources		
Use of data from outside sources		**Value**
Researcher's discipline	with a lot of time and effort for pre-processing	$X^2_{(2)} = 8.093; p = 0.017$
	I do not use data from outside sources	$X^2_{(2)} = 7.021; p = 0.030$
Researcher's country	without any problems	$X^2_{(2)} = 8.706; p = 0.013$

	with a lot of time and effort for pre-processing	$X^2_{(2)}=$**48.814**; p = 0.000
	I do not use data from outside	$X^2_{(2)} =$9.556; p = 0.008
Researcher's years of experience	with a lot of time and effort for pre-processing	$X^2_{(5)}=$18.734; p = 0.002

Nearly three-quarters (72%) of researchers feel that research data should be stored at the researchers' own universities for long-term storage. Majority (65%) feel that costs for this storage should be borne by the universities, as opposed to 53% who feel that it should be paid by the funding bodies, and 39% who feel it should be paid by a national agency (Figure 6).
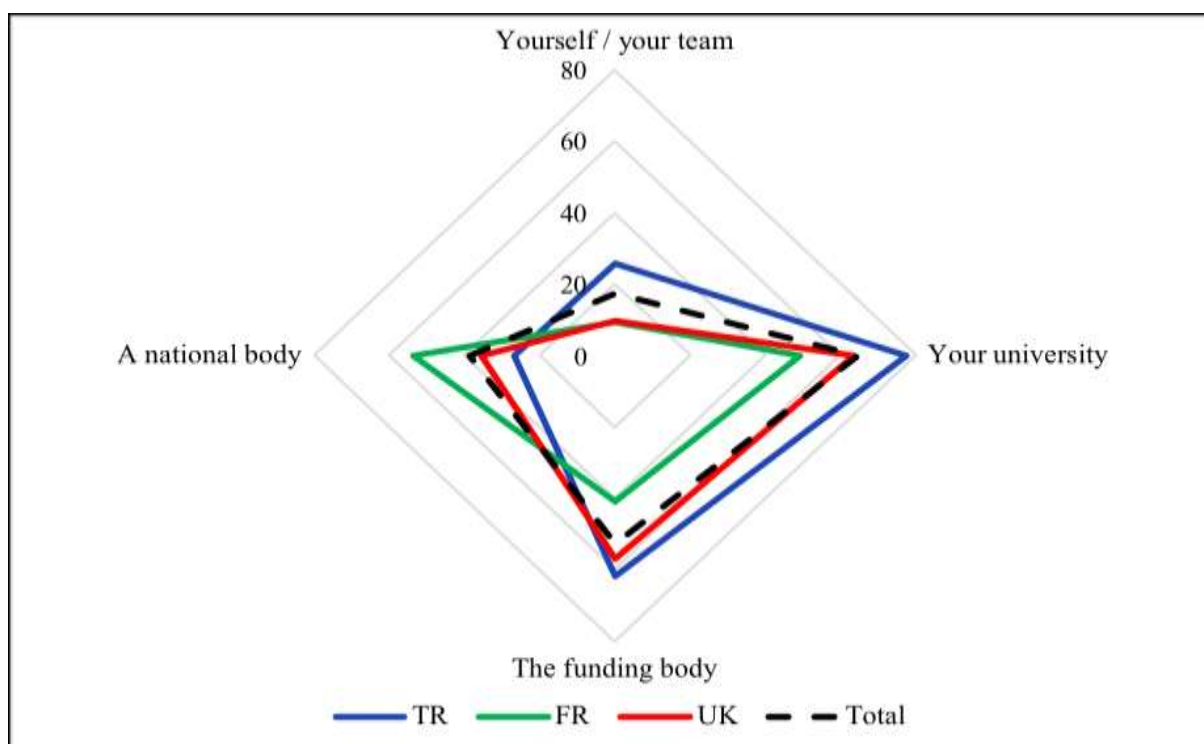

Figure 6: Expectations for funding for data storage and OA by country

There are some associations between funding body expectations and attributes such as country and discipline (Table 3). Whilst 78% of researchers in Turkey expect their university to fund data storage and public access, it is 64% in UK and 49% in France. On the other hand, expectation from a national body for funding is highest in France 54% (as opposed to 36% in UK and 27% in Turkey).

Table 3: Associations between researchers' attributes and expectations for funding for data storage and OA	
Funding body expectations	**Value**

Researchers' country	Yourself/your team	$X^2_{(2)}$=**50.179**; p = 0.000
	Your university	$X^2_{(2)}$=**86.124**; p = 0.000
	Funding body	$X^2_{(2)}$=**44.085**; p = 0.000
	National body:	$X^2_{(2)}$=**76.698**; p = 0.000
Researchers' discipline	Your university	$X^2_{(2)}$=11.801; p = 0.003

Despite a growing demand from governments and funding bodies for OA to research data, discussed earlier in the paper, less than half of the researchers (46%) are familiar with the OA requirements. Researchers have different views on the potential benefits and challenges of OA and data sharing: 56% of researchers *strongly agree* or *agree* to share research data willingly and comfortably; and about 44% of researchers foresee no problems with sharing research data. However, more concerns were noted when the same question appeared in the context of data sharing in OA mode (see Figure 10). Data ethics was also a concern: 64% *strongly agree* or *agree* that data ethics could be an issue for data sharing (Figure 7). Data indicates that some of the researchers who claim having no problems with data sharing still have some concerns regarding ethics.

A significant association was found between researchers' country and universities'

encouragement towards OA ($X^2_{(4)}$ = 30,371; *p* = 0,000). The lowest score is in France (18%) and Turkey (19%) and the highest in UK (41%). There is also a strong association between country and the researcher's familiarity with funding body's requirements for data storage ($X^2_{(4)}$ =193.026; *p* = 0,000). Researchers in the UK are more familiar with funding body's requirements (67%); these scores are very low for France and Turkey (11% and 16% respectively). These findings could be linked with the development of RDM related policies and mandates in the respective countries.
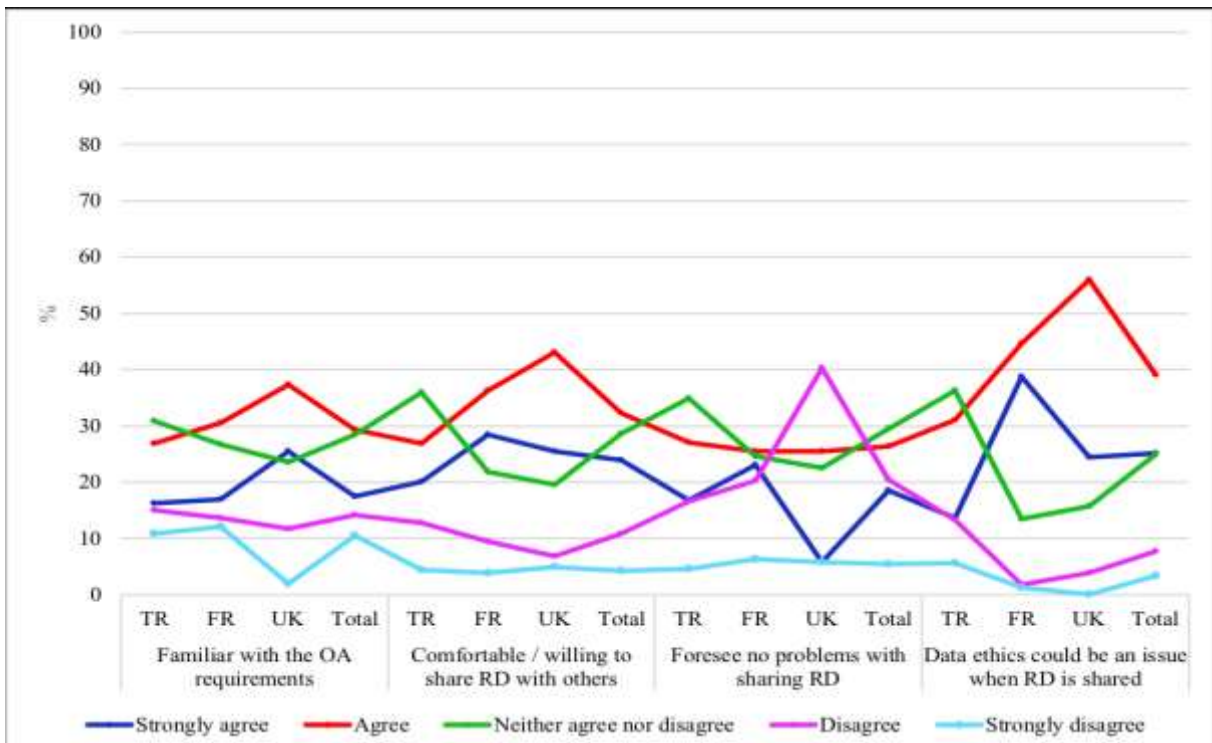
Figure 7: Familiarity with and views on OA by country

The culture of OA is not very common: only 27% of researchers share their data with everyone, and majority prefer some form of a restricted access; 35% open data only to their own research team, and 46% are willing to make data available only on request (Figure 8-9).
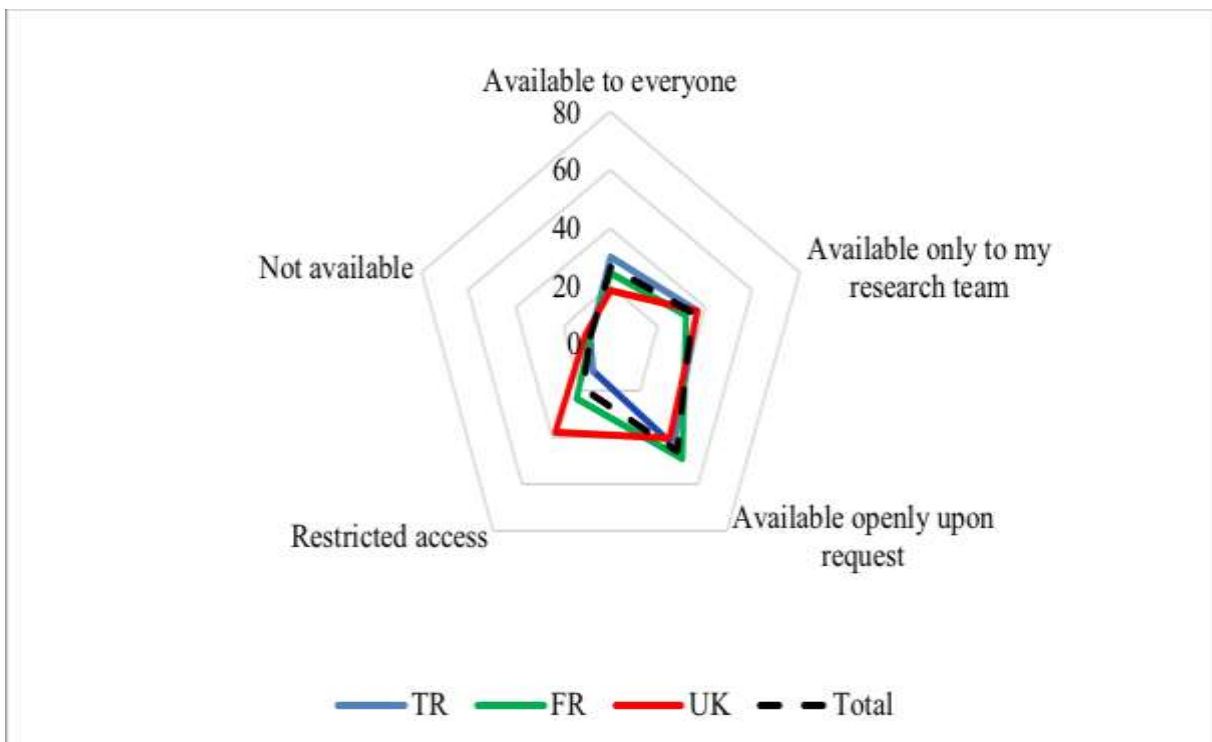

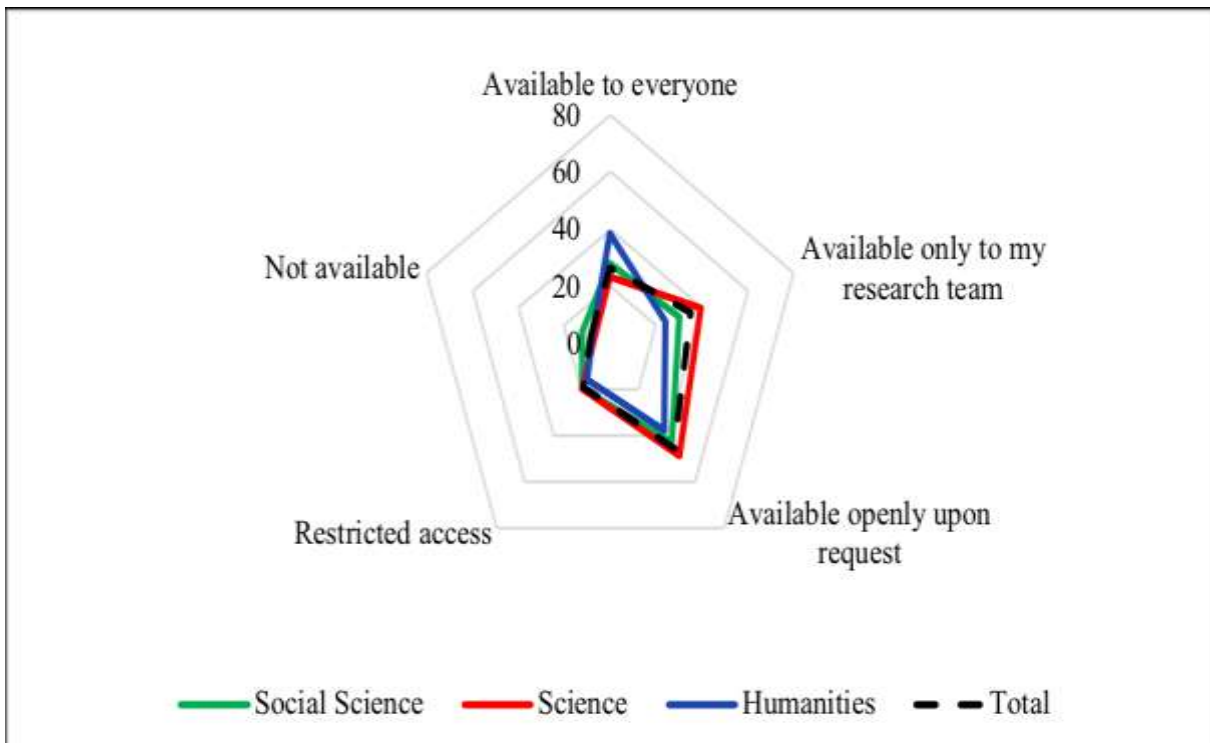Figure 8: Availability of researchers' own data by country

Figure 9: Availability of researchers' own data by discipline

Researchers do have a number of concerns for making data available in OA mode. The greatest concerns are legal and ethical issues (39%), whilst other key concerns include misinterpretation of data (29%) and misuse of data (25%) (Figure 10).
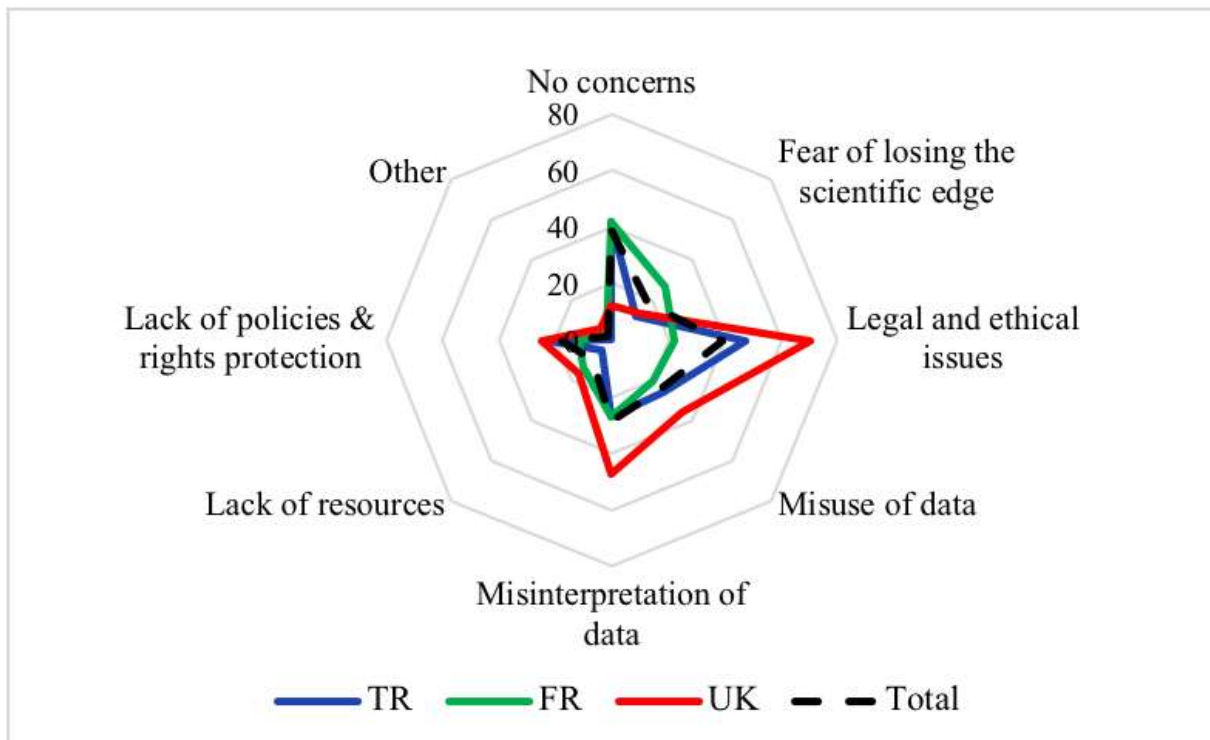

Figure 10: Researchers' concerns for sharing their data by country

Table 4 shows associations between researchers' attributes and behaviour in relation to opening their data for public access. Findings shown in Figure 10 indicate that concerns regarding legal and ethical issues is highest in the UK (70%, as opposed to 47% in Turkey and 22% in France); and number of researchers who have no concerns is the lowest in the UK (13%), as opposed to France and Turkey (42% each); fear of losing scientific edge is higher in France (27%). Data indicates that well established policies increase the concerns of researchers probably because of the effect on their awareness level and mandates. Additionally, concerns regarding legal and ethical issues are the highest in social sciences (49%).

Table 4: Associations between researchers' attributes and their concerns on sharing data		
Concerns for sharing data		**Value**
Researchers' country	Legal and ethical issues	$X^2_{(2)}=$**110.779**; p = 0.000
	No concerns	$X^2_{(2)}=33.870$; p = 0.000
	Fear of the losing scientific edge	$X^2_{(2)}=$**37.124**, p = 0.000
Researchers' discipline	Legal and ethical issues	$X^2_{(2)}=18.497$; p = 0.000
	No concerns	$X^2_{(2)} =8.525$; p= 0.014
	Fear of the losing scientific edge	$X^2_{(2)} = 8.060$; p = 0.018
Researchers' years of experience	Legal and ethical issues	$X^2_{(5)}=16.603$; p = 0.005
	Lack of resources	$X^2_{(5)}=17.983$; p = 0.003

Metadata and tagging of datasets

Figures 11 and 12 show the types of coding or tagging used by researchers on their datasets. Nearly one third of researchers (29%) do not assign tags to their datasets. Half of the researchers

(50%) assign administrative information, 35% assign discovery information, and only 24% assign technical information for tagging.
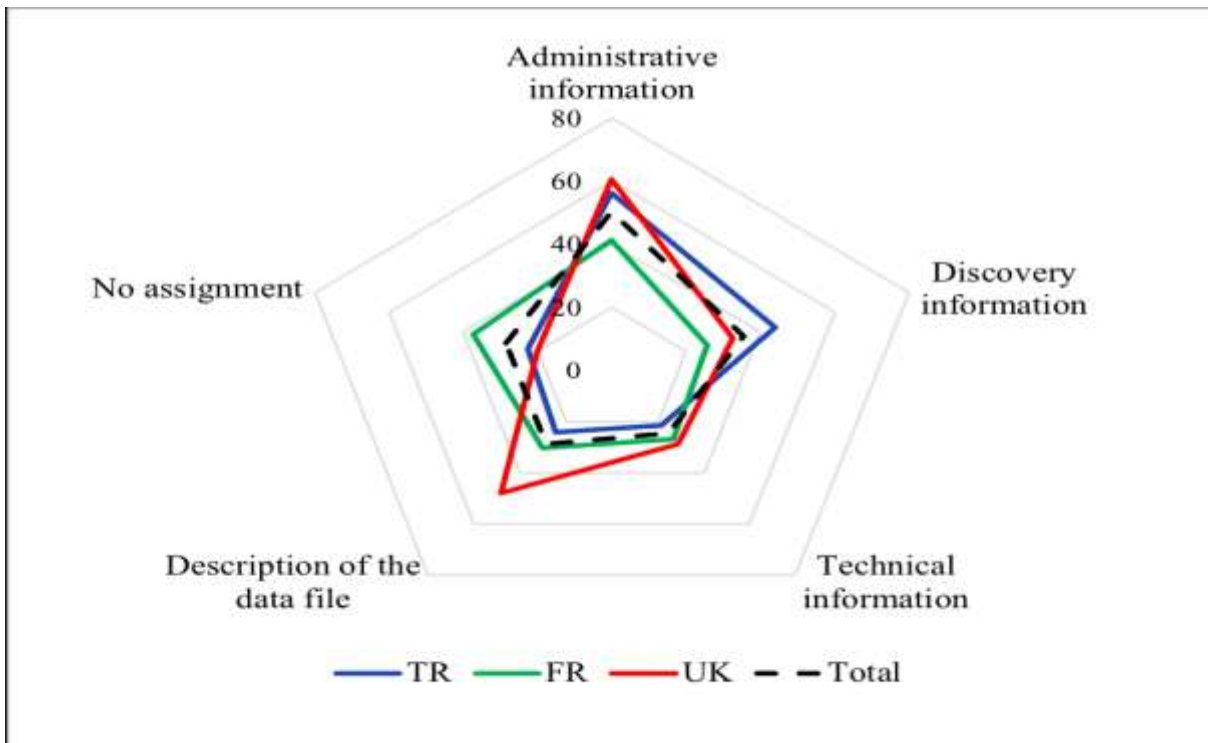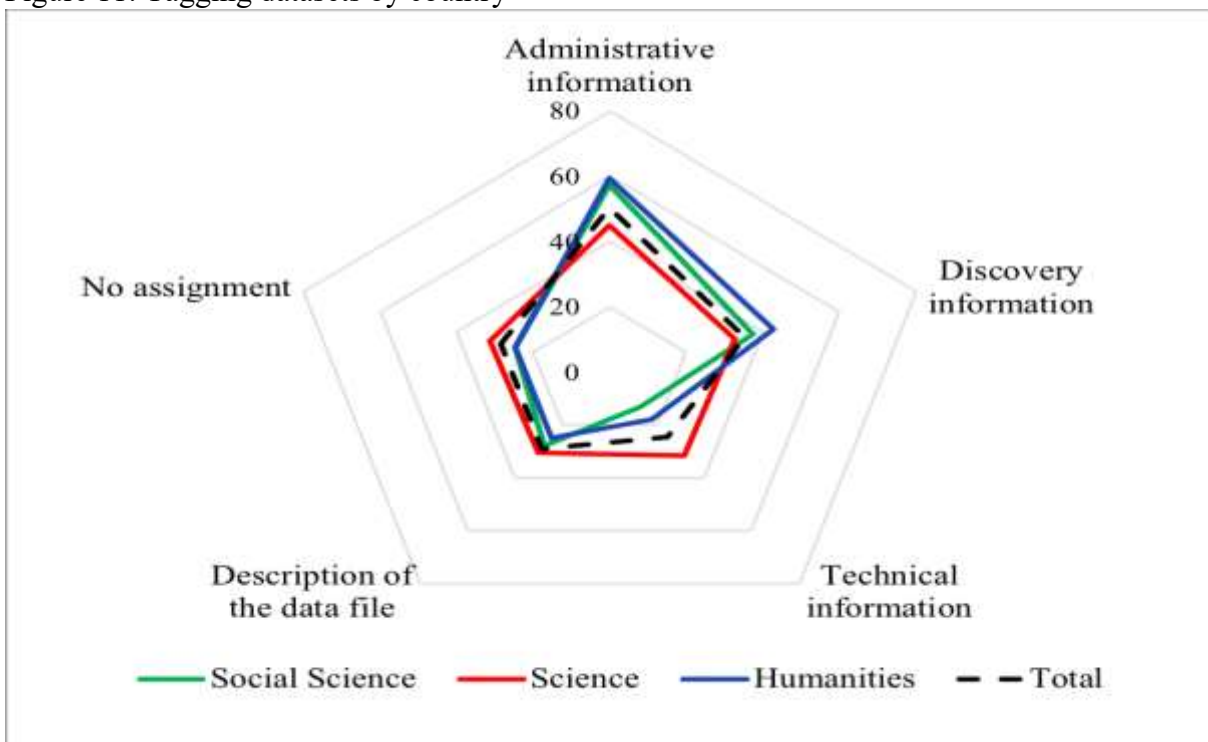
Figure 11: Tagging datasets by country

Figure 12: Tagging datasets by discipline

Some associations were detected between the use of metadata and researchers' attributes (Table 5). The number of researchers who do not assign tags and metadata to their datasets is higher in France (37%); additionally, assigning administrative (41%), and discovery (26%) information to datasets is the lowest in France. As for the discipline, using technical tags is higher in sciences (32%), while the use of administrative tags is higher in humanities (59%) and social sciences (57%).

| Table 5: Associations between researchers' attributes and tagging datasets		
Tagging datasets		**Value**
Researchers' discipline	Administrative information	$X^2_{(2)}$ =18.347; p = 0.000
	Discovery information	$X^2_{(2)}$ =6.335; p = 0.042
	Technical Information	$X^2_{(2)}$ =**43.766**; p = 0.000
Researchers' country	Administrative information	$X^2_{(2)}$ =26.032; p = 0.000
	Discovery information	$X^2_{(2)}$ =**35.784**; p = 0.000
	Description of the data	$X^2_{(2)}$ =24.600; p = 0.000
	No I don't assign	$X^2_{(2)}$ =30.592; p= 0.000
	Lack of resources	$X^2_{(5)}$=17.983; p = 0.003
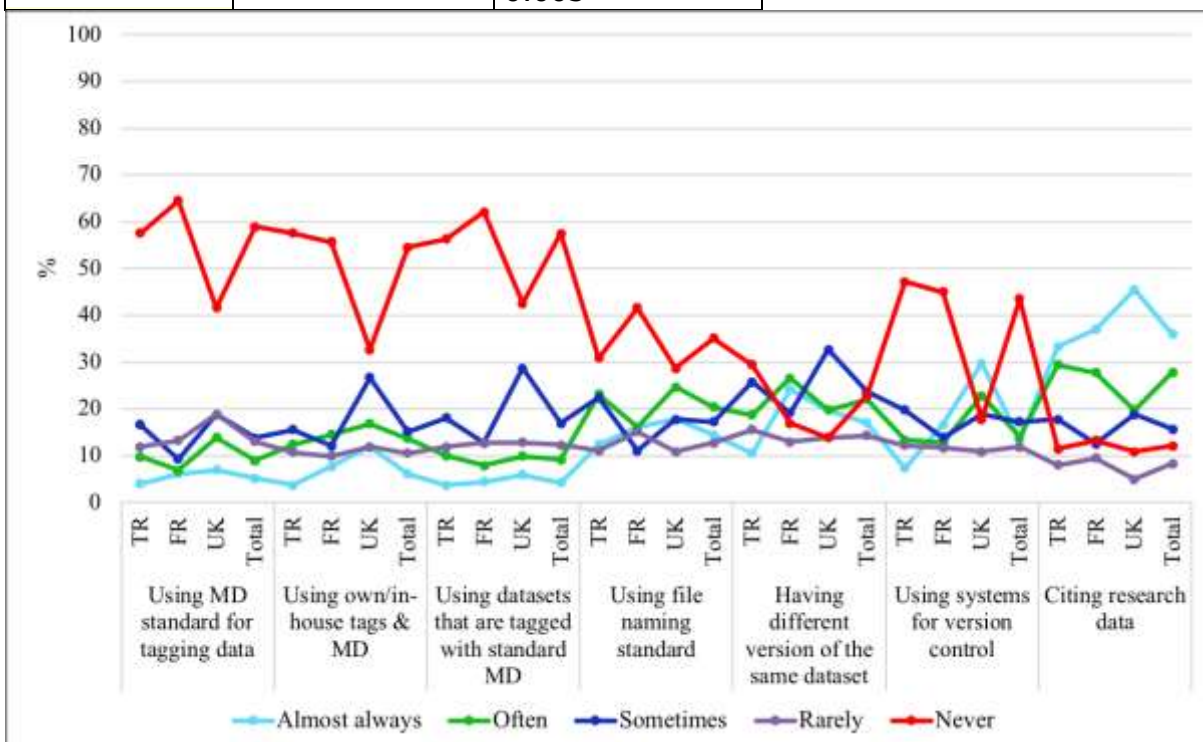
Figure 13: Metadata and file naming practices by country

Although 71% of researchers tag their data, only 14% claimed that they either almost always or often use a metadata standard for tagging data, 20% claimed they *almost always* or *often* use in- house tags and metadata, and about 72% of researchers never or rarely use a metadata standard. Data also show that very few researchers practise or use standard file naming systems, a key requirement of a good DMP. Almost half of the researchers (48%) rarely or never use a file naming standard. More than half

(51%) claim that their research community does not use a standard for file naming while 34% are uncertain about this. More than half of researchers (58%) are uncertain whether their university has a standard file naming system whilst 36% claim that their university does not have such a standard; 39% of researchers always or often create different versions of the same datasets, and only 27% claim using systems/techniques for version control to easily recognise a specific version of their data (Figure 13).

More than half (55%) of the researchers are either uncertain or not familiar with the concept of metadata. Nearly 92% of researchers are either uncertain or do not know whether their university has a prescribed metadata set for uploading data onto the repository. Almost half (44%) of researchers reported using a standard data citation style, but only 31% of universities have a recommended citation style; 60% are familiar with the concept of DOI, but only 35% of the researchers have a unique researcher ID. A majority (61%) of researchers in UK own a unique researcher ID, while this is only 27% in Turkey. Whilst approximately 45% of researchers in both the UK and Turkey claim that their universities recommend some guidelines for citing data, for France it is only 12%. About half (45%) of the researchers from France also claimed that they do not use a standard style for citing research data (Figure 14).
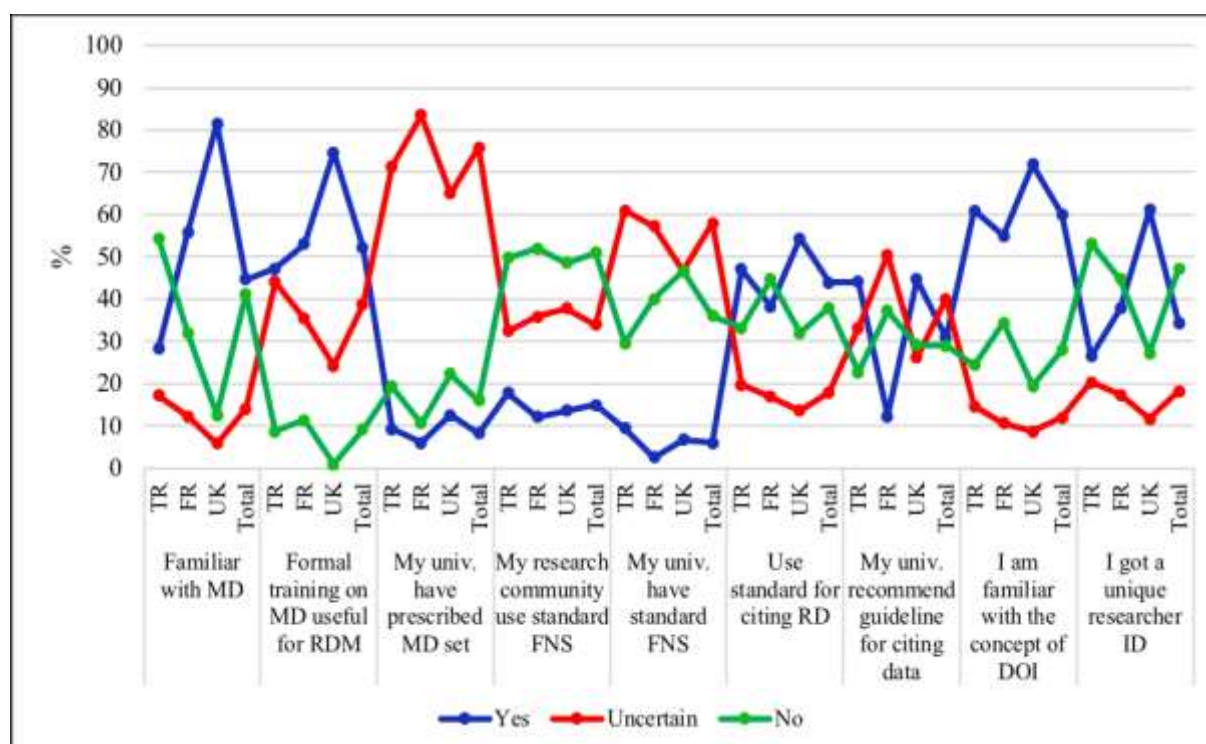

Figure 14: Familiarity with metadata (MD), citing and file naming standards (FNS), DOI, researchers' ID by country

Table 6 shows the association between researchers' attributes and behaviour in relation to data tagging and metadata. Use of metadata standards is highest in

France (65%) as opposed to Turkey (58%) and UK (42%). Use of in house metadata standard is lowest in Turkey (almost always: 4%, never: 58%). Turkey had the lowest score for familiarity with metadata (28%). UK researchers are more in favour of a metadata training (UK: 75%, France: 53%, Turkey: 47%).

Table 6: Associations between researchers' attributes and metadata practices		
Statistically significant association with		**Value**
Researchers' country	Using metadata standard	$X^2_{(8)}$=30.944; p = 0.000
	Using own/in house metadata standard	$X^2_{(8)}$=36.136; p = 0.000
	Using datasets tagged with standard metadata	$X^2_{(8)}$=21.665; p = 0.006
Researchers' country	Familiarity with metadata	$X^2_{(4)}$ = **138.636**; p = 0.000
	Formal training on metadata would be useful	$X^2_{(4)}$ = 33.795; p = 0.000
	University have metadata set	$X^2_{(4)}$=26.994; p = 0.000

Data management plans: issues and awareness

Despite various government and funding body mandates, researchers still appear to be unfamiliar with DMP: 75% of researchers are uncertain whether their institution has a DMP, and only 15% of researchers have DMP for their current research, and only 13% have used a DMP for their past research. Only 29% of researchers believe that a DMP helps researchers manage their data (Figure 15).
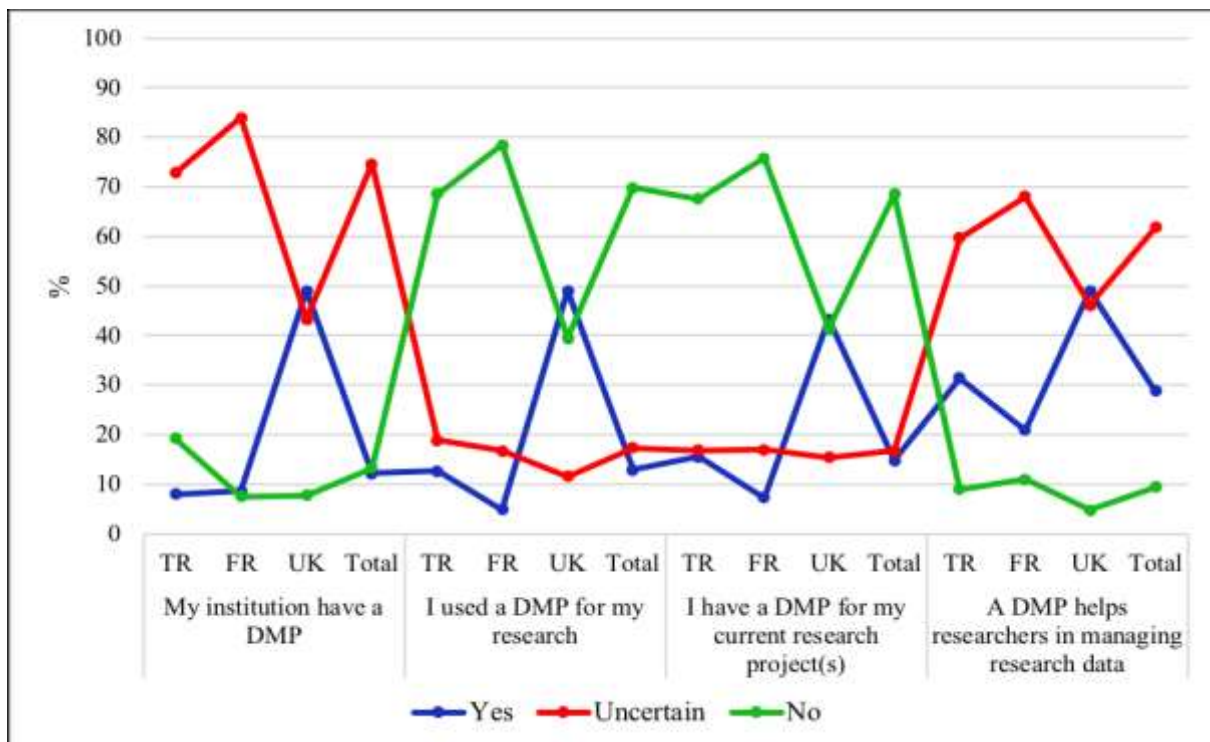
Figure 15: Availability of and awareness about DMP by country

It is interesting to note that 86% of researchers either *strongly agree* or *agree* that data should be stored beyond the lifetime of a project. 73% of researchers either *strongly agree* or *agree* that every university should have a DMP; and 64% either *strongly agree* or *agree* that every university should have a prescribed metadata set for uploading data into a repository (Figure 16).
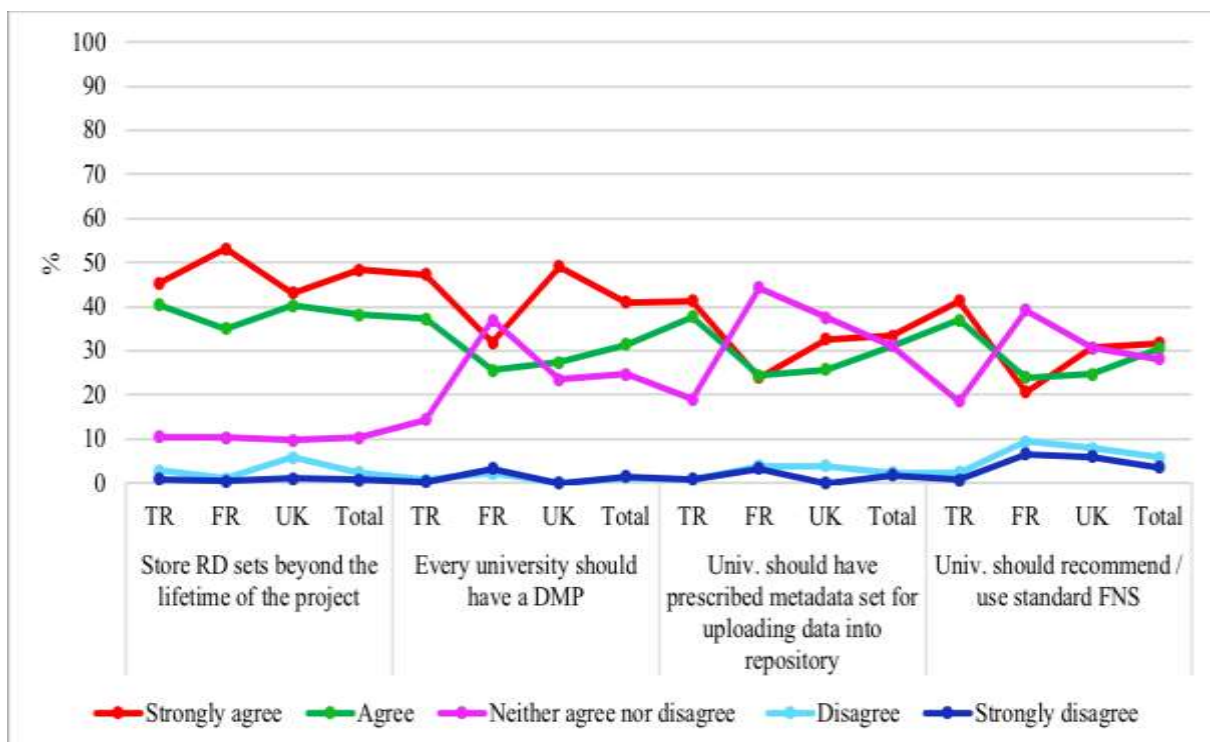

Figure 16: Opinions about RDM and DMP by country

Associations between researchers' attributes and behaviour in relation to DMP appear in Table 7. The number of researchers who believes that DMP helps in managing research data is higher in UK (49% as opposed to 32% in Turkey and 21% in France) as well as the number of researchers who have a DMP for their current research project (43% as opposed to Turkey 16% and France 7%). Percentage of researchers who have used a DMP and claim that their institution has got a DMP are the highest in UK (both 49% as opposed to <10% in the other two countries).

Table 7: Researchers' attri butes and opinions, av ailability and awareness of a DMP		
Opinions, availability and awareness of a DMP		**Value**
Researchers' country	DMP helps in managing research data	$X^2_{(4)}$=37.087; p = 0.000
	I have a DMP for my current research project	$X^2_{(4)}$ = 89.689; p = 0.000
	I used DMP for my research	$X^2_{(4)}$=150.292; p = 0.000
	My Institution have a DMP	$X^2_{(4)}$ = 177.022; p = 0.000
Researchers' discipline	Universities should recommend and use a standard FNS	$X^2_{(8)}$ = 41.219; p = 0.000
	Every university should have a prescribed metadata set for uploading data into a repository	$X^2_{(8)}$ = 21.192; p = 0.007
Researchers' country	Universities should recommend and use a standard FNS	$X^2_{(8)}$ = 132.242; p= 0.000
	Every university should have a prescribed metadata set for uploading data into a repository	$X^2_{(8)}$ = 108.751; p = 0.000
	Every university should have a DMP	$X^2_{(8)}$ = 99.024; p = 0.000

Training requirements

Findings in Figure 17 show that 81% of researchers did not have any formal training in RDM. However, it is good to note that many researchers are interested in receiving formal training in various RDM issues, e.g. 56% in DMP, 50% in metadata, 40% in version control of datasets, 44% in consistent file naming and 38% in data citation styles (Figure 18).
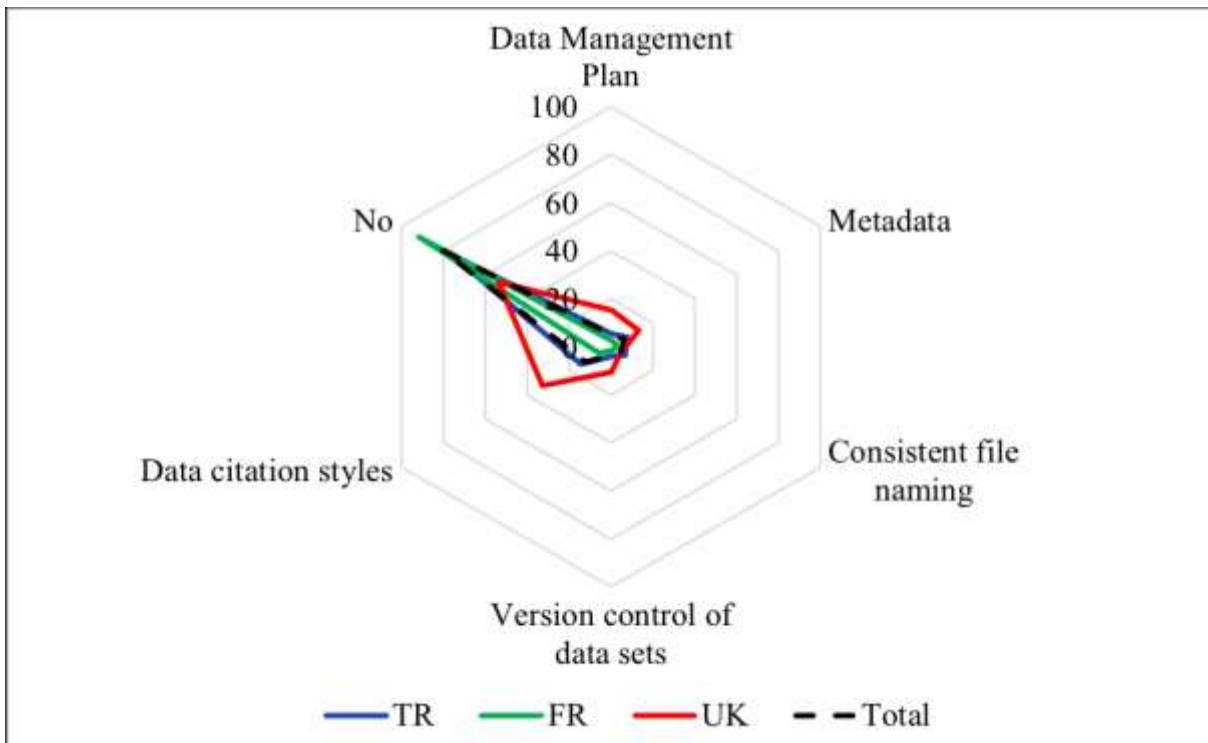
Figure 17: Familiarity with metadata (MD), citing and file naming standards (FNS), DOI, researchers' ID by country
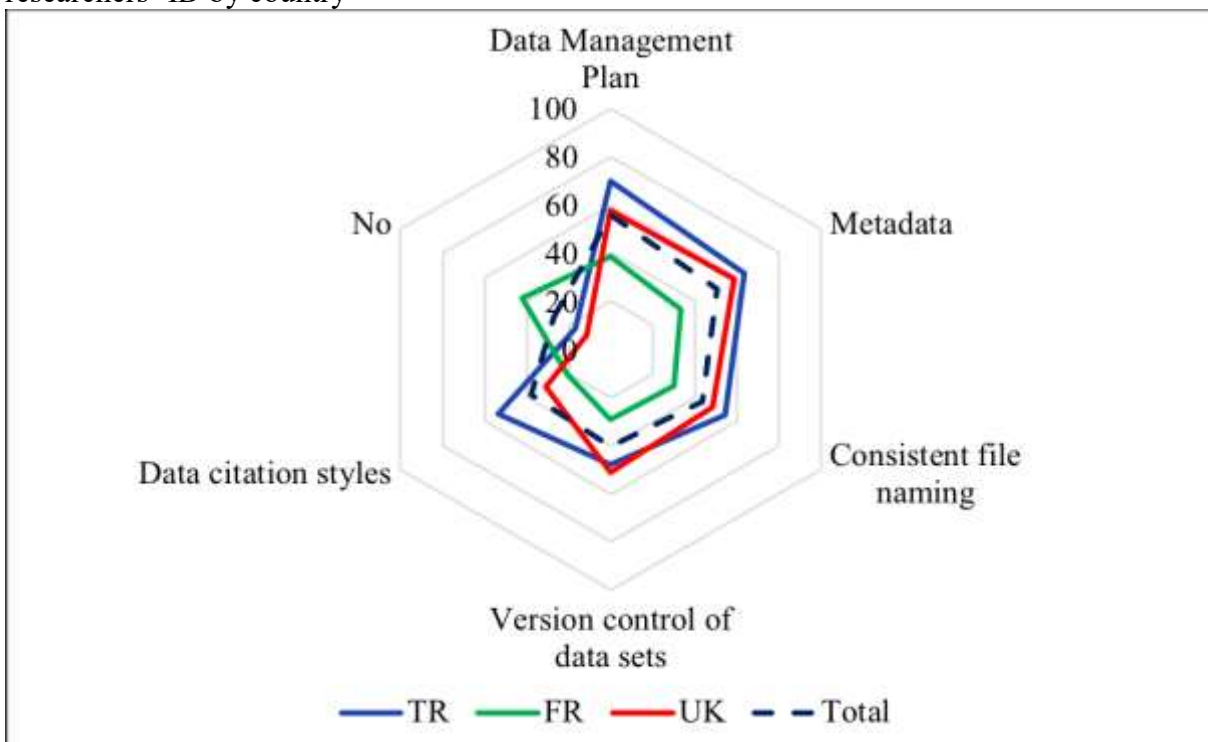
Figure 18: Training required by country

Associations between researchers' attributes and experience/opinions in relation to RDM training appear in Table 8. There is a very low association between training received and the discipline. Some aspects of training such as training on data citation styles is higher in UK (33% as opposed to 6% in France and 14% in Turkey). Associations between willingness to have training and researchers' discipline is quite

low while it is higher for country. For instance, willingness for training is generally highest in Turkey (generally more than half) and lowest in France. Moreover, researchers who have no interest in training is highest in France (43% as opposed to Turkey: 17%, and UK: 12%).

Table 8: Associations between researchers' attributes and RDM training		
RDM training		**Value**
Researchers' discipline	Have training: DMP	$X^2_{(2)}$ =8.433; p = 0.015
	Have training: Metadata	$X^2_{(2)}$ =6.566; p = 0.038
	Have training: Data citation styles	$X^2_{(2)}$ =23.188; p = 0.000
	No, I haven't had training	$X^2_{(2)}$ =28.217; p = 0.000
Researchers' country	Have training: DMP	$X^2_{(2)}$ =29.532; p = 0.000
	Have training: Metadata	$X^2_{(2)}$ =21.901; p = 0.000
	Have training: Consistent file naming	$X^2_{(2)}$ =14.717; p = 0.001
	Have training: Version control of dataset	$X^2_{(2)}$ =18.632; p = 0.000
	Have training: Data citation style	$X^2_{(2)}$ =**61.062**; p = 0.000
	No, I haven't had training	$X^2_{(2)}$ =**89.008**; p = 0.000
Researchers' discipline	Willingness for training: DMP	$X^2_{(2)}$ =10.811; p = 0.004
	Willingness for training: Metadata	$X^2_{(2)}$ =7.488; p = 0.024
	Willingness for training: Consistent file naming	$X^2_{(2)}$ =10.646; p = 0.005
	Willingness for training: Version control of dataset	$X^2_{(2)}$ =7.874; p = 0.020
	Willingness for training: Data citation styles	$X^2_{(2)}$ =9.219; p = 0.010
	No, I am not interested	$X^2_{(2)}$ =10.158; p = 0.006
Researchers' country	Willingness for training: DMP	$X^2_{(2)}$ =**97.923**; p = 0.000
	Willingness for training: Metadata	$X^2_{(2)}$ =**93.700**; p = 0.000
	Willingness for training: Consistent file naming	$X^2_{(2)}$ =**58.200**; p = 0.000
	Willingness for training: Version control of dataset	$X^2_{(2)}$ =**41.816**; p = 0.000

	Willingness for training: Data citation style	$X^2_{(2)}$ =**113.813**; p = 0.000
	No, I am not interested	$X^2_{(2)}$ =**93.908**; p = 0.000

Discussion

The key findings of this study support those of several previous studies mentioned earlier in the paper; for example,

- data sourcing, creation and use and researchers' sharing behaviour (Borgman, 2012; Mayernik *et al.*, 2013; Wallis *et al.*, 2013);
- major data sharing concerns such as trust and, ethics (The data harvest..., 2014; Faniel *et al.*, 2013; Faniel *et al.*, 2012; Yakel and Faniel 2014);
- the need for data management training for researchers (Carlson *et al.*, 2011; Wallis *et al.*, 2013; Koltay, 2015);
- the need for an increased awareness regarding the requirements for data sharing in OA mode (Carlson *et al.*, 2011) and
- training needs in metadata and tagging (Borgman *et al.*, 2012; Koltay, 2015; MacMillan, 2014).

However, our study provides some specific details in terms of researchers' behaviour on specific aspects of RDM:

- Most researchers tend to share data within the same team, but such sharing becomes limited or with restrictions when it involves sharing data with everyone.
- OA to research data is not common yet, and researchers have a number of concerns around OA.
- Legal and ethical issues appear to be the major concern irrespective of the researchers' country or discipline, while other key concerns include misinterpretation and misuse of data.
- Very few researchers appear to be using standard metadata for tagging their dataset. Using technical tags is higher in sciences (32%), while the use of administrative tags is higher in humanities (59%) and social sciences (57%). Further qualitative studies can find answers to these behavioural differences.
- Only 14% of researchers either almost always or often use a metadata standard for tagging data, and about 72% of researchers never or rarely use a metadata standard. Most researchers are not familiar with metadata, and don't know whether their institutions have a

prescribed metadata set. This could be due to a lack of proper training and advocacy.

- Researchers in the UK are more favourable towards the usefulness of metadata training (UK: 75%, France: 53%, Turkey: 47%). This may be due to the government and funding body mandates for DMP for all research projects.
- Some significant differences were noted in terms of researchers' behaviour in relation to DMP and their country. For example, nearly half of UK researchers know about, and use, a DMP for their research while these figures are very low for researchers from France and Turkey. These point towards training and advocacy issues.
- Most (81%) researchers have not had any training in RDM.
- Most researchers believe that having a DMP would be helpful, and there is a very low association between training received and discipline. In other words, researchers in every discipline lacks training in RDM and DMP.

Conclusion

UK Data Archive (2016) recommends some activities in the research data lifecycle such as data creation, data access, analysis and re-use, are undertaken or primarily driven, by researchers, while the others – data management and preservation – form part of the key professional activities of data managers. Overall, our study shows some clear gaps in awareness and understanding. This indicates an overall a lack of researcher capability in managing and sharing research data. Such gaps are evidenced by e.g., poor familiarity with OA requirements, metadata and tagging, DMP, file naming systems and version control, DOI and unique researcher ID. Together these demonstrate the need for more research and training in all aspects of RDM. Some associations between researchers' disciplines and their behaviours with regard to RDM practices were noted, but these need to be explored further at specific discipline levels rather than at broad discipline category level as done in this study. Similarly further research is needed to establish researchers' country and their behaviour with regard to RDM practices by choosing representative study samples from many more countries.

About the author

Yurdagül Ünal, Associate Professor at the Department of Information Management, Hacettepe University, Beytepe, 06800, Ankara, Turkey. She received her PhD from Hacettepe University, and her research interests are in bibliometrics, research data management, and collection development. She can be contacted at unal.yurdagul@gmail.com

Gobinda Chowdhury, Professor and Head, iSchool, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. He received his PhD from the University of Sheffield, and his research interests are in information access and users, research data management, and sustainable information services. He can be contacted at gobinda.chowdhury@northumbria.ac.uk

Serap Kurbanoglu, Professor at the Department of Information Management, Hacettepe University, Beytepe, 06800, Ankara, Turkey. She received her PhD from the University of Sheffield and her research interests are in information literacy, information behaviour, research data management, and green libraries. She can be contacted at kurbanogluserap@gmail.com

Joumana Boustany, Associate Professor at DICEN IDF EA 7339 Université Paris-Est Marne-la-Vallée, IFIS 5-7 cours du Danube 77700 Serris, France. She received her PhD from the Université Bordeaux-Montaigne and her research interests are in information literacy, open data and data management. She can be contacted at jboustany@gmail.com

Geoff Walton, Senior Lecturer in the Department of Languages, Information & Communications, Manchester Metropolitan University, Manchester, M15 6LL, UK. He received his PhD from Loughborough University and his research interests are in information and digital literacy, information behaviour, and e-learning. He can be contacted at g.walton@mmu.ac.uk

References

- Aydınoğlu, A.U. (2014). *The social, technical, and policy landscape of in environmental sciences in Turkey*. Paper presented at the 5th International Symposium on Information Management in a Changing World at Antalya, Turkey.

- Aydınoğlu, A.U. (2016). *Araştırma verileri yönetimi: Türkiye. [Research data management: Turkey]*.Paper presented at the 5th National Open Access Conference in Ankara, Turkey.

- Aydınoğlu, A.U., Doğan, G. & Taşkın, Z. (2017). Research data management in Turkey: perceptions and *Library Hi-Tech*, *35(2)*, 271-289.

- Ball, A. & Duke, M. (2015). How to track the impact of research data with metrics . Retrieved from Digital Curation Centre Web site: http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_Track_pdf

- Beagrie, N. & Houghton, J. (2013). The value and impact of data sharing and curation: a synthesis of three recent studies of UK research data centres. Retrieved from JISC Web Site:http://repository.jisc.ac.uk/5568/1/iDF308_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

- Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology, 63(6)*, 1059–1078.
- Borgman, C.L. (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge, MA, The MIT Press.
- Borgman, C.L., Wallis, J.C. & Enyedy, N. (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal of Digital Libraries, 7(1-2)*, 17-30.
- Borgman, C.L., Wallis, J.C., & Mayernik, M.S. (2012). Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work, 21(6)*, 485-523.
- Carlson, J., Fosmire, M., Miller, C.C. & Nelson, M.S. (2011). Determining data information literacy needs: a study of students and research faculty. *portal: Libraries and the Academy, 11*(2), 629-657.
- Cartier, A., Moysan, M. & Reymonet, N. (2015). *Réaliser un plan de gestion de données*. Paris: Université Paris Descartes; Université Paris Diderot; Sorbonne Paris Cité. Retrieved from http://cache.media.education.gouv.fr/file/Comment_Participer/36/0/realiser_un_dmp_406360.pdf.
- [Concordat on open research data](#) (2015). Version 10, 17th July, 2015. Retrieved from http://www.rcuk.ac.uk/documents/documents/concordatopenresearchdata-pdf/
- *[The data harvest: how sharing research data can yield knowledge, jobs and growth](#)*. *An RDA Europe report* (2014). Retrieved from https://rd-alliance.org/sites/default/files/attachment/The%20Data%20Harvest%20Final.pdf
- European Commission. (2016). *[H2020: guidelines on fair data management in Horizon 2020](#)*. V3. 3, July Programme. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Faniel, I.M., Kriesberg, A. & Yakel, E. (2012). [Data reuse and sensemaking among novice social](#)*Proceedings of the American Society for Information Science and Technology, 49*(1), 1-10. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/meet.14504901068/pdf
- Faniel, I., Kansa, E., Whitcher Kansa, S., Barrera-Gomez, J. & Yakel, E. (2013). The challenges of digging data: a study of context in archaeological data reuse. In *Proceedings of the 13th ACM/IEEE-CS*

joint conference on Digital libraries (JCDL '13) (pp. 295-304). New York, NY:

- Frank, E.P. & Pharo, N. (2016). Academic librarians in data information literacy instruction: a case study in meteorology. *College & Research Libraries*, *77(4)*, 536-52.

- Gürdal, G. & Bitri, E. (2015). *Araştırma verisi yönetimi, açık veri ve Avrupa Birliği bilimsel veri altyapısı: OpenAIRE2020* [Research data management, open data and European Union scientific data infrastucture]. Akademik Bilişim Konferansı, XVII. Retrieved from http://ab.org.tr/ab15/ozet/124.html

- Koltay, T. (2015). Data literacy: in search of a name and identity. *Journal of Documentation*, *71(2)*, 401-

- MacMillan, D. (2014). Data sharing and discovery: what librarians need to know. *The Journal of Academic Librarianship, 40(5)*, 541-549.

- Mayernik, M.S., Wallis, J.C. & Borgman, C.L. (2013). Unearthing the infrastructure: humans and sensors in field-based scientific research. *Computer Supported Cooperative Work*, *22*(1), 65-101.

- OECD (2007). Principles and Guidelines for Access to Research Data from Public Funding. Retrieved from https://www.oecd.org/sti/sci-tech/38500813.pdf

- Research Councils UK (2015). *Common principles on data policy*. Retrieved from https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/

- Strasser, C. (2015). *Research data management: a primer publication of the National Information Standards organization*. Baltimore, MD: NISO.

- Tonta, Y. (2013). Açık erişimin geleceği ve araştırma verilerine açık erişim. [Future of open access and open access to research data]. Paper presented at Library Seminars at Bilkent University, 17 December 2013, Ankara. Retrieved from http://library.bilkent.edu.tr/activities/librarianship-seminars/presentations/yasar-tonta.pptx

- UK Data Archive (2016). *Create & manage data. Research data lifecycle. Concordat on open access data* *(2015)*. Version 10, 17th July, 2015. Retrieved from http://www.oulu.fi/yliopisto/sites/default/files/content/ConcordatOpenResearchData.pdf

- Uhlir, P.F. & Cohen, D. (2011). Internal document. Board on Research Data and Information, Policy and Global Affairs Division, National Academy of Sciences.

- Yakel, E. & Faniel, I. (2014). *Virtuous circles: circulating old data through new collaborations*. Paper presented at 17th ACM Conference on Computer Supported Cooperative Work and Social Computing Workshop: Sharing, Re-Use and Circulation of Resources in Cooperative Scientific Work. Baltimore, MD, February 15, 2014.
- Wallis, J.C., Rolando, E. & Borgman, C.L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE 8*(7),